

# FILTERING INTERLOPERS FROM GALAXY SURVEYS

KAZE WONG

The Chinese University of Hong Kong, Shatin, NT, Hong Kong SAR, The People's Republic of China  
 McWilliams Center for Cosmology, Department of Physics, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA, 15213,  
 U.S.A.

ANTHONY R. PULLEN

McWilliams Center for Cosmology, Department of Physics, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA, 15213,  
 U.S.A.

SHIRLEY HO

McWilliams Center for Cosmology, Department of Physics, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA, 15213,  
 U.S.A.

*Draft version June 30, 2016*

## Abstract

We present Intercut, a Python-based program that applies secondary line identification and photometric cuts to mock galaxy surveys, in order to simulate interloper identification. This program can be used to optimize the removal of interloper contamination in upcoming surveys. Intercut reads a mock galaxy survey and an emission line sensitivity and simulates interloper removal through secondary line identification and broad-band photometry. This program is designed to use the COSMOS mock catalog, although the program can be modified for an alternative mock catalog. The output of the program returns an interloper fraction for each emission line, as well as the total fraction over all lines, as a function of redshift. We test Intercut by predicting interloper rates for the WFIRST emission line sensitivity, finding agreement with previous work. This program is publically available on Github<sup>a</sup>.

## 1. INTRODUCTION

A new era of large galaxy surveys is on the horizon. Upcoming surveys like Euclid (Laureijs et al. 2011) and the Wide-Field InfraRed Survey Telescope (WFIRST) (Spergel et al. 2013), each of which comprise over a billion galaxies, will obtain new constraints on dark energy by mapping the large-scale structure (LSS) of our universe up to redshift  $z \sim 2$  (ESA/SRE 2011). Since the redshift range of these surveys is extended to a higher limit than previous ones, the signal to noise ratio (SNR) and the number of emission lines in each galaxy's spectrum is expected to be lower compared to the Sloan Digital Sky Survey (SDSS) (York et al. 2000).

Since fewer lines are observed, it will be more difficult to confirm the identity of a particular emission line. This can lead to misidentifying interloping emission lines as the survey's intended one, producing an incorrect estimation of the galaxy's redshift. These interlopers will bias the measured matter power spectrum, along with the cosmological parameters. For example, it was shown in Pullen et al. (2015) that an interloper rate of 0.15% would significantly bias the WFIRST measurement of the growth rate, a gravity probe within LSS.

Two simple methods can help distinguish interlopers from emission line galaxies, namely secondary line identification and photometric cuts (Kirby et al. 2007). Although more complex methods for removing interlopers (Newman 2008; Ménard et al. 2013) exist, they are limited by survey parameters, while the simple methods

can be applied more generally and to each individual galaxy. Recently, Pullen et al. (2015) used these methods to predict an interloper fraction of 0.2% for the WFIRST H $\alpha$  survey, and up to 10% for the WFIRST OIII survey.

In this paper, we introduce Intercut, a program that applies secondary line identification and photometric cuts to a mock galaxy survey, in order to predict the potential interloper fraction for any upcoming survey. The mock catalog used in the test run is the COSMOS Mock Catalog (CMC) (Ilbert et al. 2009; Jouvel et al. 2011). Aside from the catalog, an emission line sensitivity file containing information on the noise of the spectrograph as a function of wavelength, the noise for each photometric band, and the photometric cuts used is needed.

The Intercut workflow is as follows: First, Intercut reads data from the catalog's FITS file, then a calibration of the galaxy fluxes using a user-defined luminosity function will be executed if needed. Next, emission line and photometric band fluxes are perturbed. Finally, secondary line identification and photometric cuts are applied, and the resulting interloper fractions are computed. We test Intercut by computing interloper fractions assuming WFIRST's sensitivity, finding it agrees with previous work. In addition to the individual interloper fractions for the each line, the total interloper fraction over all lines is calculated.

The paper is structured as follows: In section 2, we gives a brief review of the interlopers' effects on the matter power spectrum and cosmological parameters. In section 3, three major functions of the program are

<sup>a</sup> <https://github.com/kazewong/Intercut>

presented: calibrating luminosity functions, perturbing emission lines and photometric band fluxes, and identifying interlopers. In section 4, we present sample output for WFIRST.

## 2. INTERLOPER REVIEW

We define the interloper fraction as in Pullen et al. (2015),

$$f(\lambda_{SEL} - \lambda_{Int}, z) = \frac{N_{Int}(z_{Int})}{N_{SELG}(z_{SELG}) + N_{Int}(z_{Int})}, \quad (1)$$

In Eq.(1),  $N_{SELG}$  represents the number of survey emission line galaxies at redshift  $z_{SELG}$ , and  $N_{Int}$  represents the number of interlopers at redshift  $z_{Int}$ . By equating the observed wavelengths of the survey and interloper emission lines,  $z_{Int}$  in fact depends on  $z_{SELG}$ :

$$\lambda_{obs} = \lambda_{rest}(z + 1) \quad (2)$$

$$\lambda_{Int}(z_{Int} + 1) = \lambda_{SELG}(z_{SELG} + 1)$$

$$z_{Int} = \frac{\lambda_{SELG}(z_{SELG} + 1)}{\lambda_{Int}} - 1, \quad (3)$$

Surveys can also have multiple interloping lines; the expression defined in Eq. (1) is an expression for only one interloper, including all types of interlopers in the survey. An expression for the interloper fraction for one line relative to the total number of galaxies including all interlopers is

$$f_i = \frac{N_{Int,i}(z_{Int,i})}{N_{SELG}(z_{SELG}) + \sum_i N_{Int,i}(z_{Int,i})} \quad (4)$$

where the sum is over all interloping lines, and the total interloping fraction including interlopers from all interloping lines is

$$f_{tot} = \sum_i f_i. \quad (5)$$

The expression in Eq. (1) can be applied directly to a single interloping line for examining performance of secondary line identification and photometry.

An interloper's effect on the matter power spectrum is given in Pullen et al. (2015)) as:

$$P_{Int}(f | k, \mu, z_{SELG}) = (1 + f_{tot})^2 P_{SELG}(k, \mu, z_{SELG})$$

$$+ \sum_i f_i^2 \gamma_{\perp}^2 \gamma_{\parallel} P_{Int,i}[q(k, \mu), \mu_q(\mu), z_{Int,i}], \quad (6)$$

where  $\mathbf{x}$  represents the observed position assuming the interloping galaxy has a redshift  $z_{SELG}$ , and  $\mathbf{y}$  is the real position of the galaxy. Note that we alter the expression to include the effect of multiple interloping lines, and  $\gamma_{\parallel}$  and  $\gamma_{\perp}$  are defined as,

$$(\mathbf{x}_{\perp}, \mathbf{x}_{\parallel}) = (\gamma_{\perp} \mathbf{y}_{\perp}, \gamma_{\parallel} \mathbf{y}_{\parallel})$$

$$\gamma_{\perp} = \frac{D(z_{SELG})}{D(z_{Int})}$$

$$\gamma_{\parallel} = \frac{\lambda_{Int} H(z_{Int})}{\lambda_{SEL} H(z_{SELG})}. \quad (7)$$

A more comprehensive discussion along with examples of power spectrum and growth rate measurement biases due to interlopers can be found in Pullen et al. (2015).

## 3. PROGRAM DETAIL

### 3.1. COSMOS Mock Catalog

The COSMOS Mock Catalog (CMC) (Ilbert et al. 2009; Jouvel et al. 2011) predicts the emission line strengths of 564,555 high-redshift galaxies. Data read in the program can be categorised into three types: galaxy properties, band properties, and line properties. Galaxy properties are information inherent to the galaxies such as redshift, half-light radius, and distance modulus. Band properties includes the magnitude, magnitude depth and noise of a band, and line properties includes flux, noise and interloper fraction of a line, which are read into two different classes and passed around in the program.

### 3.2. CALIBRATION OF LUMINOSITY FUNCTION

As luminosity functions are updated over time, it is necessary to update corresponding emission line fluxes in the mock catalog. Recall the Press-Schechter form of the luminosity function,

$$n(x; \phi^*, \alpha) dx = \phi^* x^{\alpha} e^{-x} dx, \quad \text{where } x = \frac{L}{L^*}, \quad (8)$$

Since the luminosity function is not monotonic, it is difficult to obtain the calibrated luminosity by inverting the new luminosity function, so the cumulative luminosity function is used instead of the luminosity function itself,

$$n(> x; \phi^*, \alpha) = \int_x^{\infty} \phi^* x^{\alpha} e^{-x} dx = \phi^* \Gamma(x; \alpha + 1) \quad (9)$$

To obtain the correct flux for a galaxy's emission line, the original cumulative luminosity function value needs to be calculated first, which gives the number density of galaxies with a luminosity greater than the given luminosity value. The calibrated luminosity is then computed by finding the inverse of the new cumulative luminosity function from the same number density. In other words,

$$n_{new} = n_{old} \quad (10)$$

$$L_{new} = n^{-1}(n_{new}; \phi^* x^{\alpha}) \quad (11)$$

Note that  $n$  without any subscript is the old luminosity function.

We perform this procedure to recalibrate H $\alpha$  and OIII emission lines in the CMC. Table 1 shows the luminosity functions used in the CMC and the default luminosity functions used here for calibration. Users can provide their own luminosity functions for calibration when needed. To specify a user-defined luminosity function, parameters  $(\Phi^*, L^*, \alpha)$  must be provided for each redshift bin.

### 3.3. PERTURBATION IN FLUX AND PHOTOMETRY

Obtaining the noise for emission lines is straightforward: given a sensitivity file containing the emission

TABLE 1

THE LUMINOSITY FUNCTION FOR  $H\alpha$  FROM GEACH ET AL. (2010) IS THE LUMINOSITY FUNCTION USED IN THE CMC, AND THE LUMINOSITY FUNCTIONS FOR  $H\alpha$  AND OIII FROM COLBERT ET AL. (2013), THE LUMINOSITY FUNCTIONS USED FOR CALIBRATING THE  $H\alpha$  LINE AND OIII DOUBLET. FOR REDSHIFTS OUTSIDE THE RANGE OF SPECIFIED LUMINOSITY FUNCTION PARAMETERS, THE PARAMETERS ARE SET AT THE VALUE OF THE CLOSEST BOUNDARY, INSTEAD OF EXTRAPOLATION.

| source             | line      | redshift range        | luminosity function parameters  |
|--------------------|-----------|-----------------------|---|
| Geach et al 2010   | $H\alpha$ | $z < 1.3$             | $\Phi^* = 1.37 \times 10^{-3}, \alpha = -1.35, L^* = 5.1(1+z)^{3.1} \times 10^{41}$ |
|                    |           | $z \geq 1.3$          | $\Phi^* = 1.37 \times 10^{-3}, \alpha = -1.35, L^* = 6.8 \times 10^{42}$            |
| Colbert et al 2013 | $H\alpha$ | $0.3 \leq z < 0.9$    | $\Phi^* = 10^{-2.51}, \alpha = -1.27, L^* = 10^{41.72}$                             |
|                    |           | $0.9 \leq z \leq 1.5$ | $\Phi^* = 10^{-2.7}, \alpha = -1.43, L^* = 10^{42.18}$                              |
| Colbert et al 2013 | OIII      | $0.7 \leq z < 1.5$    | $\Phi^* = 10^{-3.28}, \alpha = -1.5, L^* = 10^{42.39}$                              |
|                    |           | $1.5 \leq z \leq 2.3$ | $\Phi^* = 10^{-3.60}, \alpha = -1.5, L^* = 10^{42.83}$                              |

line noise as a function of observed wavelength, the program interpolates the noise for a particular line. If half-light radii are also included in the sensitivity file, a 2 dimensional interpolation can be performed.

We perturb photometric band flux according to the photometric error. In order to obtain the photometric error for a galaxy with finite radius, we use the magnitude depth given by:

$$m_{dep} = m_{dep,0} - 1.25 \log_{10} \left( 1 + \left( \frac{r}{r_0} \right)^2 \right), \quad (12)$$

Where  $m_{dep,0}$  is the magnitude depth for a point source, and  $r_0$  is the characteristic radial size of the point-spread function for that band.

The program supports perturbing the spectral lines' flux and magnitude in photometric bands, which can give an estimation of the interloper fraction uncertainty, given by the standard deviation of the resulting interloper fractions for each simulation. The perturbation in flux is straightforward; for a spectral line with noise lower than a cut, the flux will be perturbed. We perturb the flux of the line using normal distribution,

$$f_{new} = f_{original} + noise * \mathcal{N}, \quad (13)$$

Since data in photometric bands is given in magnitudes, it must be transformed to a flux before being compared to a threshold:

$$f = 10^{-0.4m+48.6} \text{ erg cm}^{-2} \text{ s}^{-1} \text{ Hz}^{-1} \quad (14)$$

If the flux is higher than a threshold, then that band's flux will be perturbed in the same way as the flux for emission lines.

### 3.4. IDENTIFYING INTERLOPERS

Following the scheme in Pullen et al. (2015), we use two methods are used in the program to distinguish interlopers from survey galaxies—secondary line identification and photometric cuts. From Eq. (2), we know the ratio between two emission lines wavelengths remains the same regardless of redshift. Thus, we can use a pair of lines to rule out the redshift ambiguity of the source of an emission line. Whenever we observe a line, we can try to find another line in the spectrum with a certain separation, and if the secondary line is found, then we can confirm that galaxy is a survey emission galaxy.

For interlopers that survived the secondary line identification procedure, colour cuts are applied to distinguish them from SELGs, note that the definition of colour here is referring to the difference in different photometric bands. Since the interloping galaxy in a sufficiently small redshift range will have a redshift different from the SELGs, the two may have different colours. This can be used to identify interlopers.

After applying those cuts to the galaxy set, the leftovers are potential interlopers. Together with Eq. (5), the interloper fraction as a function of  $z_{SELG}$  is outputted as an expected result from the program.

## 4. RESULT

To examine the power of cuts, we should use the expression from Eq. (1):

$$f_{I,cut} = \frac{N_{Int,i}(z_{Int,i})}{N_{SELG}(z_{SELG}) + N_{Int,i}(z_{Int,i})} \quad (15)$$

The expression in Eq. (15) contains only information about a specific interloping line and the SEL, and is therefore independent of other interloping lines. This can be used to investigate effect of cuts applied to that line. To understanding the contribution of a particular line to the total interloper fraction among the whole survey, we can make use of Eq.(4). In Figs. 1 and 2, the left columns show the effect of cuts and the right columns show the contribution to total interloping fraction.

Since the denominator in Eq.4 depends on the interloper fraction of other interloping lines at the redshift bin, if the cutting conditions applied to other interloping lines are more effective than the interloping line being inspected at some particular redshift, the contribution to the total interloper fraction for that particular interloping line will be raised. That is the reason causing the total contribution of OII line in both surveys after applying secondary line cuts are higher than only SNR cuts in some redshift bin. The sum of total interloper fraction overall redshift should not be affected by this.

In figure 3, the total interloper fraction of the surveys are shown, which gives an estimation to the total contamination of the survey.

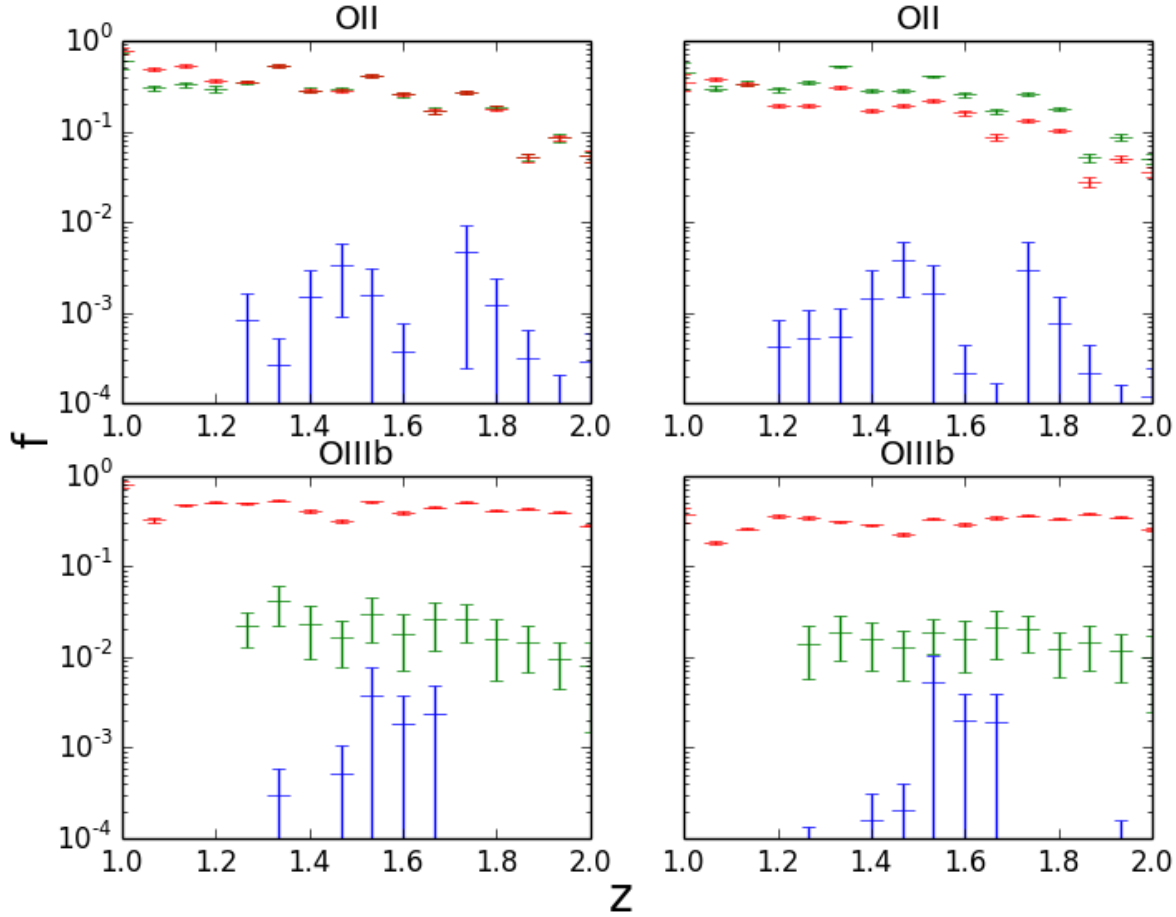


FIG. 1.— Histograms of interloper fraction for Ha survey from OII line and OIIIb line. The interloper fraction here is the interloper fraction including the interloping line and the SEL only. The binning of these histograms is  $\delta z = 0.1$ . The red line represents the interloper fraction for that line with SNR > 7 only. The green line represents the interloper fraction with SNR > 7 and secondary line identification, and the blue line represents the interloper fraction with SNR, secondary line identification, and photometric cuts. The right column is interloper fraction for that line in the whole survey, which corresponds to (4). And the left column is the interloping line and the SEL. Error comes from perturbing the flux noise in the simulation. Please note that the [OIIIb] line corresponds to [OIII 5007].

## 5. CONCLUSION

In this paper, we presented a code that estimates interloper fractions based on mock catalogues. The results obtained from this code agree with Pullen et al. (2015). For a given survey sensitivity, the interloper fraction can be determined as a function of redshift. The part that identifies interlopers according to flux and noise in emission lines and photometric bands can work on its own once the flux and noise are well defined in the program. Noise in photometric bands is now estimated based on the radial size and redshift of the object; other parameters can be added if needed. In order to modify the flux of emission lines according to

user-defined functions, a separate routine can be used. Such calibration of luminosity functions may need end-user's further implementation, and there are examples given in a separate routine on Github. Using a catalog other than the CMC may require some modifications in how the program reads catalog data.

## 6. ACKNOWLEDGEMENT

K.W. and S.H. were supported by NASA grant NASA 12-EUCLID11-0004 and 15-WFIRST15-0008 for this work. A.P. was supported by the McWilliams Fellowship of the Bruce and Astrid McWilliams Center for Cosmology.

## REFERENCES

- Colbert, J. W., Teplitz, H., Atek, H., et al. 2013, *The Astrophysical Journal*, 779, 34  
 Geach, J. E., Cimatti, A., Percival, W., et al. 2010, *MNRAS*, 402, 1330  
 Ilbert, O., Capak, P., Salvato, M., et al. 2009, *ApJ*, 690, 1236  
 Jouvel, S., Kneib, J.-P., Bernstein, G., et al. 2011, *A&A*, 532, A25  
 Kirby, E. N., Guhathakurta, P., Faber, S. M., et al. 2007, *ApJ*, 660, 62  
 Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, *ArXiv e-prints*  
 Ménard, B., Scranton, R., Schmidt, S., et al. 2013, *ArXiv e-prints*  
 Newman, J. A. 2008, *ApJ*, 684, 88  
 Pullen, A. R., Hirata, C. M., Dore, O., & Raccanelli, A. 2015, *ArXiv e-prints*

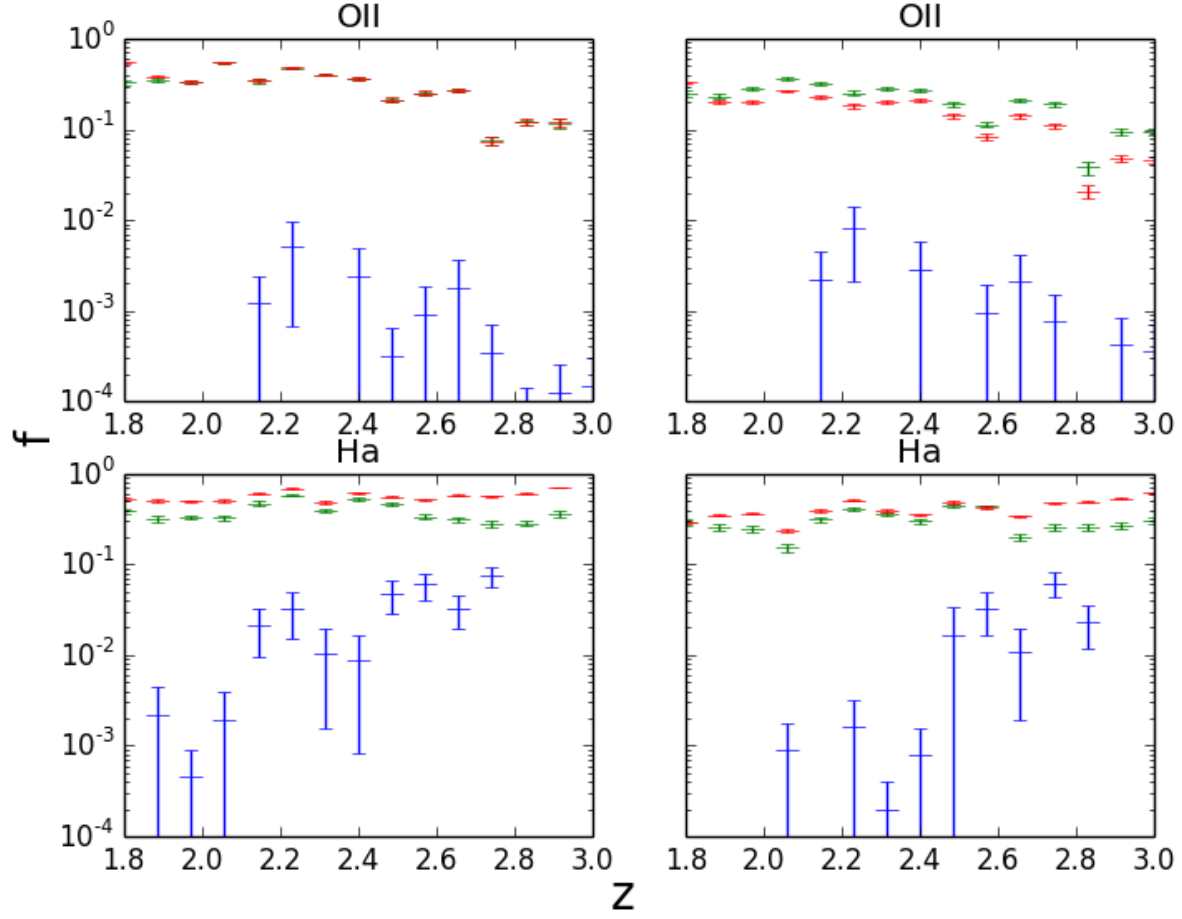


FIG. 2.— Histograms of interloper fraction for OIIb survey from OII line and Ha line, the configuration here is similar to plots in Figure 1.

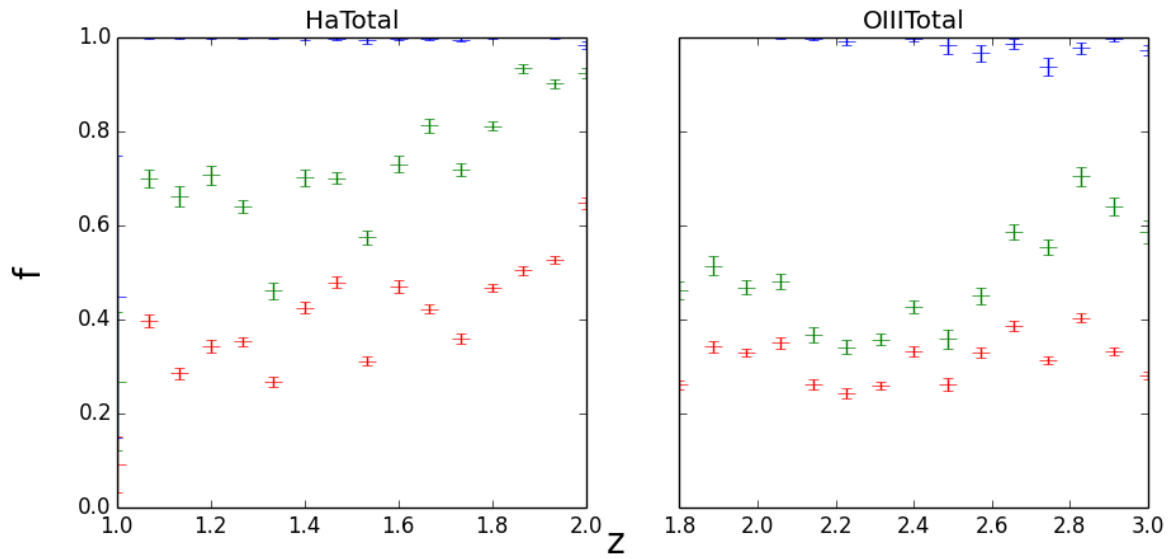


FIG. 3.— The two plots here showing the total interloper fraction in the two survey, the configuration follows previous plots.

Spergel, D., Gehrels, N., Breckinridge, J., et al. 2013, ArXiv e-prints

York, D. G., Adelman, J., Anderson, Jr., J. E., et al. 2000, AJ, 120, 1579